DOCUMENT RESUME

ED 227 652 FL 013 474

AUTHOR Bensoussan, Marsha; Ramraz, Rachel
TITLE The Fill-In Test: Combining Multiple-Choice and Modified Cloze Techniques to Test English as a Foreign Language Reading Comprehension.
PUB DATE [82]
NOTE 68p.
PUB TYPE Reports - Research/Technical (143) -- Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cloze Procedure; Comparative Analysis; *English (Second Language); Higher Education; *Language Tests; *Multiple Choice Tests; *Reading Comprehension; Reading Tests; Test Format

ABSTRACT

The fill-in-test, a modified version of the cloze procedure, is relatively easy to construct, is easier to score than the cloze procedure, and yields a greater number of questions per line of text than the conventional multiple choice comprehension test. The fill-in-test is constructed by inserting blank spaces into a text and providing a number of alternative responses for filling in each blank. Statistical analysis comparing the fill-in test with multiple choice and cloze tests showed that the three formats yield similar results. Four comparisons were conducted of these test formats with Israeli university students and applicants. The fill-in test is intended to provide an additional multiple choice technique as part of a test battery, not to replace conventional multiple choice tests. The appendices include copies of the tests used in the study. (Author/RW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED227652

The Fill-in Test: Combining Multiple-Choice and Modified Cloze Techniques to Test English as a Foreign Language Reading Comprehension *

by

Marsha Bensoussan, Department of Foreign Languages

and

Rachel Ramraz, University Assessment and Selection Unit

University of Haifa

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Marsha Bensoussan
Rachel Ramraz

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

FL 013 494

* For his help and encouragement, the authors wish to thank Dr. Baruch Nevo, Head, Department of Psychology, University of Haifa.

The authors are grateful to Dr. Micah Leshem, Department of Psychology, University of Haifa, for his thoughtful and constructive criticism.

We also wish to thank Prof. Robert Cooper, Department of Education, The Hebrew University of Jerusalem, for his helpful comments.

The Fill-in Test: Combining Multiple-Choice and Modified Cloze Techniques to Test English as a Foreign Language Reading Comprehension

ABSTRACT

This paper describes the Fill-in Test, a modified version of the Cloze procedure, whose purpose it is to test *EFL reading comprehension. The advantages of the Fill-in Test are that it is relatively easy to construct, it is easier to mark than the Cloze, and that it yields a greater number of questions per line of text than the conventional Multiple-Choice Comprehension Test.

The Fill-in Test is constructed by inserting blank spaces into a text and giving a number of alternate responses from which the testee must choose the appropriate word(s) to fill in the gap in the text. Blanks may be inserted within a range of 7 to 15 words, depending on the items tested; they can test items on the micro-level: the word level and the sentence level, and the macro-level; the inter-sentence or paragraph level and the whole-text level.

Statistical analysis comparing the Fill-in Test with both Multiple-Choice and Cloze Tests, shows that all three test formats yield similar results. In Experiment One, three test batteries, each consisting of 1 Fill-in Test and 3 Multiple-Choice Tests, were administered to 435 first-

*EFL - English as a foreign language

year students at Haifa University. Average scores on both types of test were similar.

In Experiment Two, 4 texts were used as the basis for 8 tests: 4 Multiple-Choice Tests and 4 Fill-in Tests. The 8 tests were divided among 1,487 applicants to Haifa University. The average of the raw scores of both test versions was almost equivalent. The Multiple-Choice versions had a higher percentage of questions with high Discrimination Indices (greater than .30), whereas the Fill-in Test version yielded more test items and had a higher reliability score. The paper compares and analyzes the test items in the Multiple-Choice and Fill-in Tests. Although they may not test the same reading comprehension skills, they both require the reader to focus on a specific amount of text in order to answer a test question.

The third part of this paper examines the Fill-in and Multiple-Choice subtests of the English Entrance Examinations at the Universities of Haifa and Tel Aviv. The first year, the examination was administered to 7,499 applicants; the second year a similar examination, following the same format but consisting of different texts, was administered to 7,114 applicants. Mean scores of the Fill-in and Multiple-Choice subtests were similar. Pearson correlations between the Fill-in and Multiple-Choice subtest scores were .75 and .79 for each of the two years.

Finally, this paper compares the results of each of the three Fill-in Tests in Experiment One with the results of another test battery

which includes 57 items and consists of four subtests: Sentence Completion items, Vocabulary items, a Multiple-Choice test, and a Cloze passage. This test battery was administered to 354 applicants at Haifa University.

The purpose of the Fill-in Test is not to replace the conventional Multiple-Choice Test, but to offer an additional multiple-choice test technique. If a series of texts is to be administered as part of a test battery, it is suggested that one of them be in the format of a Fill-in Test.

The Fill-in Test: Combining Multiple-Choice and Modified Cloze Techniques to Test English as a Foreign Language Reading Comprehension

INTRODUCTION

Our task at the Haifa University Selection and Assessment Unit is to administer psychometric examinations to entering candidates who wish to be accepted by the university and study in the department(s) of their choice. The English examination is one of 6-8 subtests in these psychometric examinations. It tests thousands of university candidates[1] and either places them in appropriate English classes or exempts them from further study. We are basically interested in students' ability to read academic texts in English, and toward this end, we administer a 60-minute, 55-question, Multiple-Choice Test consisting of a number of texts with accompanying content questions and a Fill-in Test. The Fill-in Test requires the testee to choose, from among several (usually 4) possible answers, the correct word or phrase that will "fill in" the gap in the text.

The Fill-in Test was developed by the writers over a period of eight years in an attempt to improve the efficiency of the Haifa

---

[1] Most candidates have already studied seven to eight years of English in order to obtain their high school diploma.

University English Examination. The purpose of this paper is to describe the Fill-in Test and its properties, to explain how to construct it, to give some statistical evidence for its effectiveness, and to compare it with traditional Multiple-Choice and Cloze Tests.

Disadvantages of Conventional Multiple-Choice Tests

One problem with constructing multiple-choice content questions is the lack of efficiency in time and materials: a student must read a great many lines of text in order to answer relatively few questions. A paragraph of approximately 100 words will yield, on the average, no more than five questions. Since a test's reliability depends, to a great extent, on the number of questions it contains, the results of a test of 20 minutes, which contains a page of text (approximately 400 words) and about 10 questions, are most often unsatisfactory. Too much time is spent on the input (reading and answering questions), and not enough output (information about the student's reading ability) is obtained. In short, this method is somewhat inefficient.[2]

---

[2] The other subtests on the psychometric examination yield 25-40 test items for a 30-minute test session.

Another problem facing the test constructor is that a text must be accompanied by a minimum number of questions if it is to be useful. Tests are usually pre-tested on a small sample of 100 to 300 students before being administered on a large scale. In the small-scale pre-testing, unsuitable questions are weeded out and only the best questions are used. If, as sometimes happens after the elimination process, too few questions remain, the entire text cannot be used. If the text is to be salvaged, completely new questions have to be re-written and pre-tested, in the hope that the final number of questions will not fall below the minimum number.

Finally it can be argued that the questions themselves constitute an extension of the text which is added to the student's burden during the test. The test mark, then, is a composite of understanding both text and questions and does not necessarily reflect comprehension of the text alone.

One Solution: The Cloze Procedure

The appearance of the Cloze procedure[3] offered a non-multiple choice solution to the problems of efficiency and testing only the text. Using Taylor's Cloze procedure, where the student is expected

---

[3] Taylor, W.L. (1953) "Cloze procedure: A new tool for measuring readaptability," Journalism Quarterly, 30: 415-453; Oller, J.W. (1973) "Cloze tests of second language proficiency and what they measure," Language Learning, 23, 1:105-118.

to write in his own words to fill the gap in the text, it was possible to obtain one test item for every n words. Leaving a blank after every 7th word, for example, would guarantee that a text would yield at least one test item per line of text, which yields a higher number of test items per line of text than the multiple-choice test.

Further experimentation, however, showed that the placement of a blank word after every nth word did not suit our needs, and the procedure was rejected for the following reasons:

1. The spacing was too close to be comfortable and, for the short text that we were using, the thread of thought easily lost.

2. Some words that were strategically placed, reflecting a reader's sensitivity to the writer's line of argument, could not be tested if the mechanical choice of blanks was to be followed. Either the blank space would have had to be moved or the text rewritten.

3. It is assumed that the information in each item is independent of that needed to answer other items (see Klein-Braley, 1981). That is, a response to a particular blank is determined by the context surrounding that blank, and not by other blanks in the text, except only indirectly.[4] For this reason, if necessary and without harming the test as a whole, it is possible to reduce the total number of test items by removing any

---

[4] Though it is possible for the test-constructor to deliberately omit the key content words (personal communication, E.A. Levenston, Dept. of English, The Hebrew University of Jerusalem), this is not one of the purposes of the Fill-in Test.

number of blank spaces and replacing them by the original word of the text. Theoretically, if a particular blank is found to be an unsuitable test item, then it can be replaced by the original word in the text, and the student can continue reading on to the next blank. An item omitted from a Fill-in Test, reducing a test from 20 to 19 or 18 items, for example, would be unnoticeable.

In the case of the Cloze Test, however, it would not be possible to remove the unsuitable blanks at will. The unsuitable test item (i.e., blank space) would need to be re-filled by the original word in the text. The gap refilled, the number of words until the next blank space would then increase, for example, from 7 to 15.

A strict adherence to the rules would not permit the omission of any unsuitable test items. On the other hand, a more flexible attitude toward the test construction would focus on the number and placement of blank spaces in the entire passage rather than on the number of words between each blank space.

Cranney (1973), however, suggests that rational Cloze deletion would be more productive than the random Cloze. Although rational deletion would increase test construction time and result in a loss of objectivity, Greene (1965) also recommends its use. Greene (1965) successfully used a modified Cloze based on rationally selected content words (nouns, verbs, adverbs, and adjectives), with an average deletion rate of one word in twelve. He found that this procedure produced a

test equal in difficulty but better in reliability and item performance (Discrimination Indices) than the standard, randomly selected Cloze.

A non-random Cloze procedure, where the deleted items mark relationships between propositions, was developed by Levenston et al. (1982). They call this type of Cloze test "Discourse Cloze" because its deletions consist solely of cohesion markers on the macro-text level. Assuming that the ultimate purpose of reading comprehension is a correct interpretation of macro-relationships, they argue that testing proficiency in the higher-level skill (macro-level) simultaneously tests proficiency at the lower (micro-) level as well.

In explaining the purpose of the "Discourse Cloze," they provide a theoretical framework by which they explain how the student is assumed to use clues in the text to fill in the blank spaces. They have adopted the distinction made by Van Dijk (1977, 1980) between micro- and macro-text levels. The former describes individual sentences or propositions in a text. Derived from the micro-structures of a text are the macro-structures, which represent intersentential, global or whole-text meanings of a discourse. According to Levenston et al. (1982):

> The inferencing process needed for reconstructing cloze tests can rely on a range of linguistic, pragmatic and textual clues.

For each of these categories, the following definitions are provided:

> By 'linguistic' we mean cases in which the meaning and form of the missing word are clearly specified by the linguistic micro-context in which the deletion occurs.

This 'linguistic' category is further subdivided into 'syntactic' and 'semotactic':

> By "syntactic" we mean cases in which the word intended is clearly specified by the grammatical construction in which the deletion occurs. By "semotactic" we mean cases in which the meaning and lexical choice of the intended word is confirmed (or, in the case of the cloze, provided) by the interaction of that word with the meaning and occurrence of other words within the context.

The 'linguistic' context does not exceed the sentence and therefore functions on the level of micro-processing.

> We have termed the ability to replace items that require extra textual knowledge in the cloze a 'pragmatic' ability. At the most general level, pragmatic ability is the general knowledge of the world that provides the basis for forming expectations and interpreting given texts.

> A third component of reading . . . is the ability to follow the cohesive ties in a text in order to understand the interrelationship between the sentences. This ability differs from linguistic and pragmatic knowledge in that it is specific to text-processing as distinct from sentence-processing.

This textual component, which embodies comprehension at the macro-level, includes the concepts of grammatical and lexical cohesion as described in Halliday and Hasan (1976). Grammatical cohesion includes the features of anaphoric reference, substitution, ellipsis, and conjunction; lexical cohesion includes the features of repetition, hyponyms (and super-ordinates), antonyms, synonyms, and collocations (on the macro-level).

Each of these features of the textual component is exemplified in the "Discourse Cloze." We will adopt the theoretical framework

described by Levenston et al. (1982) in explaining the construction of the Fill-in Test (see below).

We had also used the modified rational open-ended (or free-response) Cloze at Haifa University with good statistical results.[5] However, it did not solve our problem completely because the marking was inefficient.[6] We were forced to return, rather, to the multiple-choice format for purely technical reasons.

Other Solutions: Multiple-Choice Modification of the Cloze Procedure

The idea of combining both Cloze and Multiple-Choice procedures in one test is not new. Ozete (1977) discusses Ronald Carver's "reading in-put test" which has two alternate responses. Guthrie (1973) introduced "the maze task" -- a Cloze passage with three alternate responses for each blank: a correct answer, a syntactic alternative, (a word syntactically but not semantically appropriate), and a lexical alternative

---

5. See Bensoussan, "A Comparison of Cloze and Multiple-Choice Reading Comprehension Tests." Report No. 57. (Haifa: University Selection and Assessment Unit, June 1981.) See below, PART THREE.

6. It was inconvenient to mark thousands of examinations by hand and to constantly re-assess standards of acceptability during the marking period. Approximately 2,000 papers were marked by 20 teachers in a large room during an 8-hour marking session. Every questionable answer was put to the whole group of markers, so that marking could be consistent for every test marker. This discussion, as well as corrections and changes that resulted from the discussion, added to the marking time.

(a word neither syntactically nor semantically suitable). Cranney (1973), Porter (1975), and Jonz (1976) discuss Cloze tests with four possibilities for every blank. Porter (1975) gives general guidelines to the test constructor on how to intuitively generate effective distractors, whereas Cranney (1973) and Jonz (1976) explain how to derive distractors from student errors on previously tested open-ended Cloze tests on the same passages. According to Cohen (1980, pp. 94-95), the addition of the multiple-choice technique to the Cloze is a major change. The student performs a different task than in the open-end Cloze.

All of these multiple-choice modifications of the Cloze are based on the random Cloze, with a blank for each $\underline{n}^{th}$ word. Cranney (1973) also experimented with random Cloze. Deleting every 10th word, he compared results of two kinds of Cloze tests based on the same texts: the free-response Cloze and the machine-scorable multiple-choice Cloze. His wrong responses for the latter were taken from the most frequent incorrect responses in the former Cloze (from a different sample of students). But his results were inconclusive.

In part, some of the multiple-choice techniques are re-introduced. The student must choose the best alternate response on the basis of his comprehension of the text. Distractors, however, and not necessarily the text, can make the test difficult (Pikulski and Pikulski 1977). In spite of this disadvantage, however, the multiple-choice Cloze is recommended because it is short, and at the same time, it has a high reliability (Jonz, 1976).

Our Solution: The Fill-in Test

We used a modified version of the Cloze: words were deleted on the basis of a rational, not a random process. The rational Cloze gives the test constructor more control over specific areas of comprehension and over the items to be tested and diagnosed. The test can be tailor-made to suit a particular group of students or teaching points.

In contrast to the "Discourse Cloze" proposed by Levenston et al. (1982) which omits only those words tapping high-level (macro-) skills, the Fill-in Test includes the entire range of possible clues on all three levels: linguistic, pragmatic, and textual.

In terms of format, we printed the text on the left page, the alternate responses on the right page (facing the text), and instructed students to write the number (1, 2, 3 or 4) of the correct answer in the appropriate place on the computerized answer sheet. Instructions, therefore, were identical for the Fill-in and Multiple-Choice subtests of the test battery. This uniformity of test instructions helped to avoid confusion.

Since our basis was not strict, random Cloze, and since our test closely resembled a Multiple-Choice Test in format, we decided to call it a Fill-in Test, so as not to confuse it with standard Cloze.

We used the procedure of inserting 20-30 blank spaces into a 300-word

text. Each blank space takes the place of a word or phrase (group of not more than three words), and for each blank space, there is a choice of 4 possible answers. The Fill-in Test modifies the Cloze procedure in three ways:

1. Possible responses are already provided in a multiple-choice format -- unlike the Cloze, which contains gaps in the text that need to be filled in.

2. Unlike the Cloze test, a blank space in a Fill-in Test can take the place of more than one word.

3. Blank spaces are placed not after every nth word,[7] but within a range of 7 to 15 words or more. Each blank space touches on the comprehension of one of these four categories:

   a. word form and/or meaning

   b. sentence meaning

   c. paragraph meaning (across sentences)

   d. whole-text meaning

It is important that the focus of the Fill-in Test items be expanded to include sentence, paragraph, and even whole-text meaning. It is assumed that throughout the test, in order to fill in a blank, the student needs to use redundancy clues present in the text.

---

[7] This possibility is advocated by Alderson (1979). Blanks do not represent a random sampling of the text; each blank space focuses on a specific semantic point, making the Fill-in a discrete point test.

The Fill-in Test is not meant to be a substitute for the Multiple-Choice Test, but rather an additional multiple-choice technique. It allows more test items to be given in less time than the Multiple-Choice Test, and it is more efficiently and easily constructed. If a series of texts is to be administered as part of a test battery, therefore, it is suggested that one of them be a Fill-in.

Fill-in Test Construction

If placement of the blank spaces is based on a linguistic examination of the text rather than randomly, it might be argued that one way of constructing the Fill-in Test would be to pre-test it as a Cloze. Likely blank spaces could be chosen by the test constructor, and the possible distractors, it might be supposed, might be found from among the students' wrong answers. (see Jonz, 1976).

Experience shows, however, that only about one-third of the test items can be obtained in this way. This method is very time consuming and yields relatively little in return.[8] The best way to construct the Fill-in Test is to decide beforehand the structures and ideas that are to be tested and to place blanks where these points are likely to be tapped.

It was assumed that a blank could tap one of two levels of reading: at the micro-level -- specific understanding of a word or

---

[8] It was suggested by Dr. Valerie Whiteson, English Department, Bar Ilan University, that a pre-test using the Cloze would be successful provided that the English proficiency of the students was high enough, that is, near native level (personal communication).

collocation where the clue(s) appeared in close proximity to the blank (one or two words before and/or after), and at the macro-level - a more general understanding of the whole test (e.g., writer's opinion, key word showing comprehension of concept and contrast/opposition), main idea of paragraph).

If we adopt the theoretical framework described by Levenston et al. (1982), we can analyze Fill-in Test A (See Appendix A) as a sample test to show the distribution of the test items according to both micro- and macro-levels of reading. Since this was a real test and not just a theoretical exemplification of an ideal Fill-in Test, many of the features in the theoretical framework did not actually appear as part of the test and will therefore not be discussed in this paper. All actual items on the test, however, will be related to the theoretical model.

TEXTUAL CLUES	test item(s)
1. linguistic (micro-)	
a. syntactic	L
b. semotactic	
1) words or phrases (distractors chosen from words already in the text)	
a) nouns	E,H,M
b) adjectives	S
c) verbs	O
d) phrases	N (F)
2) collocations	

TEXTUAL CLUES	test item(s)
c. both syntatic and semotactic	
1) parts of speech (same root, give different forms), test syntax and understanding of the relationship between words	K
2) word forms, as they affect comprehension	
a) adjective	C,D
b) verb	J
c) pronoun	A,I
d) preposition	R
3) collocations	Q,R
2. pragmatic (micro-)	H,T
3. textual (macro-) -- cohesion	
a. grammatical	
1) anaphoric reference	A,I
2) conjunction	G,P
b. lexical	
1) superordinates	D,E,M
2) collocations	D

Some test items touch on more than one of the text components, either because of the multiplicity of clues in the text or because of the possibities presented by the alternate responses (distractors).

Basic Assumptions in Choosing Blanks

The basic criterion for choosing blank spaces is that there is enough redundancy in the text so that a proficient reader could use the clues to fill in the gap with an appropriate word or expression. Moreover, in choosing the blanks, the test constructor would focus on pivotal or key words in a logical argument to see whether a student can follow the thought sequence.

Since it is more difficult to find blanks indicating a student's knowledge of whole-text level words, preference would be given to using these. Function words, such as "however" as opposed to "therefore," would be good places for blanks. Content words such as nouns, adjectives, and verbs which carry the weight of an argument (as opposed to auxiliary verbs) would also be useful, and their opposites would be included among the distractors. In this way, the test constructor may suggest alternate misleading logical thought sequences, but only one set of choices would be consistent with the writer's intentions within the text as a whole.

Other items tested could be cohesive markers such as "not only . . . but also," "either . . . or," and "on one hand . . . on the other hand." It is assumed that a student's recognition of these syntactic devices would enable him to follow the flow of an argument, and that lack of recognition would impede his comprehension.

At the moment, it is assumed that test constructors intuitively choose for omission words which would reveal understanding or misunderstanding of idioms, collocations, contextual lexical clues, redundancy

of ideas or opinions, and parallel or symmetric syntactic structures. In thinking up alternate responses, then, the test constructor would be expected to use words focusing on a particular point, either in terms of content or structure. It was found, in fact, that items focusing on a single point (i.e., in which students were to choose from among 4 adjectives) were more successful than items unclearly focused (e.g., alternate responses included 2 adjectives and 2 conjunctions, so the student does not understand if he is to look for content or form or both; another ever worse possibility: 1 adjective, 1 noun, 1 conjunction, and 1 relative pronoun). The student probably does not understand the point that is being tested, and therefore the results would not really show whether the student has understood the text. We aimed for half the blanks to represent content words (nouns, verbs, adjectives, adverbs) and the other half function words (conjunctions, prepositions, word forms). For examples, see Fill-in Test A, Appendix A. Although this weighting does not reflect the frequency of these types of words in the language, it does ensure a wide variety or sampling of items tested.

Rationale for Determining Distractors

In choosing distractors, the test constructor may make use of collocations, presenting words that could appear together and make sense in some other context. For this reason, opposites are particularly useful:

they test the student's understanding of the whole text. Conjunctions are also helpful here. A student choosing "therefore" when only "however" would fit the context may have understood a particular sentence, but certainly did not grasp how the sentence fit into the context as a whole.

For testing English as a foreign language, synonymous distractors should not be used. It is advisable to avoid distractors where the correct choice is ambiguous even for native speakers. Thus, one should also avoid asking about detailed grammatical points(e.g., the distinction between it's and its) or prepositions which may also be confused by native speakers. The Fill-in Test is essentially a test of reading comprehension, not of grammar.

Comparison Between the Fill-in Test and the Multiple-Choice Test

Having developed the basic outlines of the Fill-in Test, we needed statistical proof that it would do its job as well as the more well-known, conventional type of multiple-choice test. Two experiments were run, using two separate test batteries, which compared the statistical results of the Fill-in Test with those of the Multiple-Choice test using a computerized Item Analysis procedure.[9] This analysis yields

---

[9] Rachel Ramraz "ITANA V: Computer Program for Item Analysis," Report No. 40, (Haifa: University Selection and Assessment Unit, June 1979). See also Bensoussan, Marsha "Manual for Teachers Preparing Examinations in English as a Foreign Language for Analysis by Computer," Report No. 44, (Haifa: University Selection and Assessment Unit, Sept. 1979).

the following information:

1. Easiness Index for each item (EI): % of correct responses out of all the attempted answers per item.
2. Discrimination Index for each item: point-biserial correlation between the response and student total raw test scores. A question is considered effectively able to discriminate between good and weak students if its Discrimination Index is greater than .30 but lower than .60. A Discrimination Index greater than .40 is desirable.
3. Average of the test Scores
4. Reliability (Kuder-Richardson Formula No. 20 and Split-Half Formula)

Although shrinkage is usually minimal, when pre-testing, it is advisable to place approximately 50% more blank spaces than needed. For example, if 15 items are required for a test, 22 or 23 may be pre-tested. Afterwards, when unsuitable test items are eliminated, these gaps can either be filled with the original word, or else the whole phrase in which it appears (provided it is not a key phrase) may be eliminated. The Fill-in Test still remains intact, because each item is independent of the others (see Alderson 1979, Jonz 1976, Klein-Brady 1981), even after many of the items have been eliminated.

If the statistical results of the Fill-in measure up to those of the Multiple-Choice Test, we can assume that one can be used in place of, or in addition to, the other, according to the needs of the testing situation.

EXPERIMENT ONE

Procedures

The sample consisted of 435 first-year students taking the advanced reading course in English as a Foreign Language at Haifa University in 1973. Each student took one of three English tests consisting of 4 subtests: 1 Fill-in Test and 3 Multiple-Choice (M-C) Tests (texts accompanied by multiple-choice content questions as well as by vocabulary and reference questions).

In this first test battery, the tests appear to be approximately equivalent: English Tests 1, 2, and 3 contain 77, 73, and 79 lines of text, respectively, and consist of a total of 53, 52, and 47 questions, respectively. The proportion of Fill-in Test items to M-C test items is about even: the Fill-in items make up 53%, 54%, and 51% of each of the English tests, respectively. The number of Fill-in items, therefore, does not greatly outweigh those of the M-C test.

Results

For a description of EXPERIMENT ONE, see Table 1, "Description of 3 English Tests."

TABLE 1 DESCRIPTION OF 3 ENGLISH TESTS

Test**(N=435)	English Test 1 ($n=_{154}$)			English Test 2 ($n=_{170}$)			English Test 3 ($n=_{111}$)		
Subtest	number lines	number items	$\overline{EI}$*	number lines	number items	$\overline{EI}$*	number lines	number items	$\overline{EI}$*
(1) Fill-in	30	28	52%	27	28	56%	32	24	55%
(2) M-C A***	5	6	48%	7	7	50%	12	6	63%
(3) M-C B	18	11	54%	13	6	42%	14	7	43%
(4) M-C C	24	8	51%	26	11	43%	21	10	50%
Total	77	53		73	52		79	47	
Reliability:									
Kuder-Richardson			.828			.709			.761
Split-Half			.8669			.6613			.7963

* $\overline{EI}$ = average of the Easiness Indices

** N = number of subjects

*** The relatively large number of items on this test is a result of asking mostly word meaning and not content questions.

In terms of item easiness, Table 1 indicates that Fill-in items are on a par with those of the conventional multiple-choice tests. Although the Fill-in Test constituted little more than one-third the total number of lines of text, it contributed at least one-half of the total number of test items.

Conclusions

The Fill-in Test yields statistical results that are not very different from those of the M-C Tests. In a large battery, the Fill-in Test can therefore be used in place of one of the M-C subtests. It has the advantage of yielding more questions per line of text than the Multiple-Choice Test.

EXPERIMENT TWO

Rationale

A question arises concerning the difference between the two types of test. Although they may yield similar statistical results, one could not go as far as to say that the Fill-in and Multiple-Choice (M-C) Tests examine the same skills. Nevertheless, both test the reading comprehension of a particular text. In order to have a better basis for comparison, it was decided to take 4 texts and test each text twice, using a different format and different students each time. Each test was constructed first in the conventional multiple-choice format, and the second time, the same text was used to construct the new Fill-in format.

Procedure

A total of 1487 applicants to the first-year of studies at Haifa University were tested. Most are high school graduates who have had seven to eight years of English. At random, each student

received one text with questions.

This test battery consisted of eight tests; that is, 4 separate texts -- once as an M-C Test and once as a Fill-in Test. The texts ranged in length from 28 to 36 lines; there were 9 - 15 questions on the M-C Tests and 20 - 28 questions on the Fill-in tests. (For a sample Fill-in and Multiple-Choice text pair, see Appendix A.)

Results

A comparison of the statistical results appears in Table 2, "A Comparison Between Fill-in and M-C Test Formats."

An examination of Table 2 shows that the Fill-in yield more items. The reliability of the Fill-in Test is higher, although no significant difference was found between the mean Discrimination Indices of the two test types. The M-C Tests, however, have a higher percentage of good questions (i.e., Discrimination Index greater than .30) than the Fill-in Test: Text A (100%/86%). Surprisingly, the average of the raw scores is nevertheless almost equivalent (Text A: 62%/63%) although the standard deviation of the Fill-in Test is greater.

Since there are greater differences among the average raw scores for Texts A, B, C, and D than there are between the M-C and Fill-In versions of each text, the results would indicate that the choice of text may be more important than the format by which it is tested.[10]

---

[10] The amount of pragmatic knowledge required for comprehension was a vital factor affecting differences in students' reading, according to Levenston (personal communication).

TABLE 2 A COMPARISON BETWEEN FILL-IN AND MULTIPLE-CHOICE TESTS

	Text A		Text B		Text C		Text D	
	Fill-in	M-C	Fill-in	M-C	Fill-in	M-C	Fill-in	M-C
no. lines	30	30	28	28	30	30	36	36
no. students	73	192	204	180	201	213	213	211
no. items	21	15	28	11	24	9	20	13
no. good items	18	15	23	10	22	9	14	12
Disc. Index> .30(%)	(86%)	(100%)	(82%)	(91%)	(92%)	(100%)	(70%)	(92%)
Mean Disc. Index	.44	.48	.39	.46	.43	.48	.38	.41
Stand. Dev.	13.39	6.95	10.41	8.74	9.55	5.34	13.84	13.69
t (Fill-in/M-C)	0.979*		1.832*		1.526*		0.741*	
reliability:								
Kuder-Richardson(20)	.80	.76	.79	.63	.80	.58	.69	.62
Split-Half	.80	.78	.78	.62	.80	.48	.67	.58
Average score (%) and standard deviation	63% 4.19	62% 3.28	62% 4.93	64% 2.25	75% 4.00	72% 1.88	46% 3.54	44% 2.62

*p = not significant (> .10)

Technically, the Fill-in format is at least as advantageous as the M-C Test. The remaining question is: What does it test? Having established that the Fill-in is as good a test as the M-C Test, we now ask ourselves about the nature of its function and whether this is different from that of the M-C Test.

Textual Analysis

In order to examine the functions of each test format, we analysed each test item of Test A (both M-C and Fill-in test versions). For the

tests themselves and an analysis of the functions of the questions, see Appendix A, Table A, "An Analysis of the Test Items in M-C and Fill-in Tests A."

Breaking down the results according to our criteria for choosing blank spaces (see page 14), i.e., (a) word form/meaning, (b) sentence meaning, (c) paragraph meaning, and (d) whole-text meaning, we see that the results are similar for both M-C and Fill-in Tests. (See Table 3, "Breakdown of Items According to Amount of Text Included: Multiple-Choice and Fill-in Tests A.")

Although the M-C Test and Fill-in Test may not test the same reading comprehension skills, they both require the reader to focus on a specific amount of text in order to answer a question. This

TABLE 3 BREAKDOWN OF ITEMS ACCORDING TO AMOUNT OF TEXT INCLUDED: M-C AND FILL-IN TESTS A

Category Number of Items:	Fill-in Test A	M-C Test A
a. word meaning	4	2
b. sentence meaning	10	6
c. paragraph meaning	4	4
d. whole-text meaning	3	3
Total no. of items	21	15

area of focus can vary from one word to the extent of the entire text.

The Fill-in Test need not be restricted to comprehension of single words; it can be used to tap an understanding of the wider context, in the same proportions as the M-C Test. In both, the most frequent kind of item covered sentence meaning (6 in the M-C and 10 in the Fill-in Test). Whole-text questions are most difficult to obtain in the Fill-in, but there are few of them as well in the M-C (3 questions in each).

PART THREE

Two Test Batteries Containing Both Fill-in and Multiple-Choice Items

The Fill-in Test was compared with Multiple-Choice type questions in two additional test batteries, each consisting of Fill-in and Multiple-Choice test items according to the following format: (For reasons of test security, the test batteries have not been reproduced here.)

Multiple-Choice	Subtest	A
"	"	B
"	"	C
Fill-in	"	D
Multiple-Choice	"	E

The texts were graded from A, the easiest, to E, the most difficult. For convenience, it was decided to consider all four Multiple-Choice texts as a combined single long text when comparing results with those of the Fill-in. The first test battery contained 56 items and the second 55.

Subjects:

The test batteries were administered as the English section of the Entrance Examination to the Universities of Haifa and Tel Aviv during two consecutive years. The first year, the examination was administered to 7,499 applicants; the second year, it was administered to 7,114 applicants.

Comparison Between Fill-in and Multiple-Choice Tests

The results of the two test batteries appear in Table 4, below:

TABLE 4 A COMPARISON BETWEEN FILL-IN AND MULTIPLE-CHOICE TESTS

YEAR	NUMBER OF SUBJECTS	TOTAL NUMBER OF TEST ITEMS		FILL-IN Score (per cent)		stand. dev.	K-R Rel.*	MULTIPLE-CHOICE Score (per cent)		stand. dev.	K-R Rel. *	PEARSON CORRELATIONS BETWEEN TOTAL SCORES OF FILL-IN AND MULTIPLE-CHOICE TESTS
1980	7499	56	average	9.0	(60%)	3.98	.841	21.5	(52 %)	8.75	.903	.748
			number of items:	15 items				41 items				
1981	7114	55	average	8.5	(50%)	4.32	.815	22.0	(58 %)	8.39	.880	.789
			number of items:	17 items				38 items				

* Average Kuder-Richardson reliability for two parallel test forms.

Mean scores of the Fill-in Test were similar to those of the Multiple-Choice questions. In the first test battery, the Fill-in items were 8% easier on the average than the Multiple-Choice questions; in the second test battery, the Fill-in items were 8% more difficult on the average. In this series of tests, contrary to the findings in EXPERIMENT TWO (above), the standard deviation of the Fill-in items was less than that of the Multiple-Choice questions. Apparently, Fill-in items can be made easy or difficult by the test constructor, in the same way that Multiple-Choice questions can be graded for difficulty.

As for the Pearson correlations between total Fill-in and Multiple-Choice scores, for the first test battery it was .75, and for the second it was .79. These figures are considerably higher than those obtained in PART FOUR of this study (see below), where the Pearson Correlation ranged from .36 to .47. It must be remembered, however, that the correlations obtained in PART FOUR are based on only ten multiple-choice questions, whereas the present correlations were based on 38 and 41 multiple-choice questions, respectively.

In any case, when correlations are not exceedingly high, the reason is usually that there is a factor common to both variables, and that each variable separately yields additional information. In this instance as well, the Fill-in subtests did not correlate so

highly as to permit their substitution for the Multiple-Choice subtest. Since each subtest adds another kind of information, the statistics would indicate that for improved efficiency, both kinds of subtest should be used in a test battery.

PART FOUR

In assessing the Fill-in, we wished to compare it with other types of reading comprehension tests. Accordingly, the three Fill-in subtests described in Experiment One (above, p.22) were compared with another test battery which had been previously administered to 354 applicants to Haifa University.[11] The same students took both the Multiple-Choice/Cloze Test battery and one Fill-in subtest.

Test Battery (For a copy of test battery, see Appendix B.)

The English section of the Entrance Examination consisted of 57 test items and was 75 minutes in duration. Each of the items was selected for level of difficulty and discrimination by pre-testing a similar population at Haifa University. Students were permitted to use monolingual English dictionaries. The same test was administered in spring and summer sessions with steps taken to prevent cheating.

Within the multiple-choice framework, many types of testing exercises are possible. We used the following three multiple-choice subtests:

---

11 See Bensoussan, op.cit. Reliability (Kuder-Richardson no. 20) = .93.

A. Sentence Completion subtest, which was a test of word form and syntax, where the student chose the word(s) that best completed the sentence.

B. Vocabulary Substitution subtest, a test in which the testee was asked to find the best synonym for the underlined word in each sentence.

C. Multiple-Choice Comprehension subtest, a text accompanied by multiple-choice questions about content, syntax, vocabulary, and reference.

(Same type as above, Experiments One and Two)

D. The Cloze subtest consisted of a 313-word text containing a blank space after each seventh word. Each student response was marked either correct or incorrect according to whether it was clear that the student understood the meaning of the context. Spelling errors were not counted. A panel of 12 teachers graded the examinations, and acceptable responses had to be agreed upon unanimously during the marking of papers. It was assumed that someone who was able to fill in the gaps in the text demonstrated the ability to read and understand the passage.

Sentence Completion and Vocabulary Substitution, the first two subtests, were short, consisting of only one or two sentences, whereas the Multiple-Choice Comprehension and Cloze subtests presented much longer and more complex reading passages.

Hierarchy of Difficulty

If we compare the Fill-in with each of the subtests in the test battery in terms of their respective difficulty levels (see Table 5), we obtain the following hierarchy: Vocabulary is the easiest subtest. It is followed by the more difficult Fill-in, Cloze, Multiple-Choice, and Sentence Completion subtests, all three of which are approximately of equal difficulty. This general pattern appears for each of the three Fill-in subtests.

The Cloze has the highest standard deviation, indicating a wider range of grades. It also comprises, together with the Fill-in, the greatest number of questions (26 Cloze + 28 or 24 Fill-in). The Fill-in subtests, however, have much smaller standard deviations, possibly because they were relatively easy.

TABLE 5 MEAN DIFFICULTY LEVEL FOR EACH SUBTEST

FILL-IN SUBTEST	N	Fill-in Mean	Fill-in SD	Sent. Comp. Mean	Sent. Comp. SD	Vocabulary Mean	Vocabulary SD	M-C Text Mean	M-C Text SD	Cloze Mean	Cloze SD
1	119	14.5	4.4	5.8	2.8	5.5	2.5	4.5	2.0	13.2	6.7
2	142	14.3	3.1	5.9	3.0	5.7	2.4	4.6	2.4	13.9	6.1
3	93	13.1	3.1	5.6	2.9	5.5	2.4	4.4	2.3	13.4	6.6
Total-across all 3 subtests:	354	14.1	3.6	5.8	2.9	5.6	2.4	4.5	2.2	13.5	6.4
% Mean		54%		48%		62%		50%		52%	
no. items in each subtest (100%) Fill-in subtest form:		1) 28 2) 28 3) 24		12		9		10		26	

Correlations Between Subtests

In general, the subtests do not correlate highly with each other (See Table 6), although all correlations obtained were significant. The highest correlation (.64) was found between the Sentence Completion Subtest and the Multiple-Choice text. The relatively low correlations would indicate a relation with some larger common factor, such as English reading comprehension. The subtests do not appear so close as to overlap, where one could substitute for another. In this respect, each subtest appears to be tapping a different area of reading comprehension.

It is especially interesting to note that the Fill-in subtests, multiple-choice versions of the modified Cloze procedure, do not correlate highly with either the Multiple-Choice text or the random Cloze passage.

Examining each subtest of the test battery separately, we notice the greatest variation in correlations with the Cloze passage. A relatively large range in correlations was also found between the Sentence Completion and the Vocabulary subtests. This fluctuation would indicate, on the one hand, that the Cloze procedure may not be as reliable a measure of reading comprehension as the other subtests. On the other hand, the correlation between the Multiple-Choice and Sentence Completion subtests were stable and consistently high.

TABLE 6 CORRELATIONS BETWEEN SUBTESTS

SUBTEST	Fill-in	Sentence Completion	Vocabulary	Multiple-Choice	Cloze
Fill-in					
***TOTAL	1.00	.50	.40	.41	.43
Subtest 1		.50	.38	.36	.38
Subtest 2		.53	.38	.47	.50
Subtest 3		.50	.46	.45	.48
Sentence Completion					
TOTAL		1.00	.43	.64	.54
Subtest 1			.39	.62	.54
Subtest 2			.50	.66	.60
Subtest 3			.35*	.64	.48
Vocabulary					
TOTAL			1.00	.34	.40
Subtest 1				.27**	.41
Subtest 2				.38	.42
Subtest 3				.35*	.36
Multiple-Choice					
TOTAL				1.00	.53
Subtest 1					.52
Subtest 2					.64
Subtest 3					.39
Cloze					
TOTAL					1.00
Subtest 1					
Subtest 2					
Subtest 3					

All correlations are significant, and all are p = .0001 except:

* p = .0006 and ** p = .0029.

*** TOTAL: across all three Fill-in subtests.

IN CONCLUSION

The purpose of the Fill-in Test is not to replace the conventional M-C Test but to offer an additional multiple-choice test format. The Fill-in tests reading comprehension -- not only words and word forms at the micro-level, but, more importantly, the ability to follow a logical thought sequence at the macro-level of reading. Each blank should test one specific point -- either structural or ideational -- in the text. Such a point may be found in a single sentence or may span the entire text. This wide range of focus does not occur in current multiple-choice versions of the Cloze test which are commercially available (See, for example, the Michigan Test of Language Proficiency.[12])

Statistically, the Fill-in Test measures up to the traditional M-C test. A test constructed with the Fill-in format will probably have items of the same average difficulty and effectiveness as if it had been constructed in the conventional M-C format. The only differences are that it will probably have more test items, and therefore the reliability will be slightly higher. The Discrimination Indices, however, would probably be lower.

---

12 English Language Institute, University of Michigan.

The Fill-in Test has some other advantages: it is easier to construct than the M-C Test. Each distractor tests only one point and contains only one word or short phrase. Moreover, after pretesting to eliminate unsuitable test items, the Fill-in Test is more likely to remain intact than the M-C Test, for which it is sometimes difficult to find enough test items.

For these reasons, we use a Fill-in subtest in our English examination. By permitting the inclusion of a greater number of test items in the same amount of time as one M-C Test, the Fill-in Test helps increase the reliability and improve the quality of the test battery as a whole.

References

Alderson, J. Charles. "The Cloze Procedure and Proficiency in English as a Foreign Language," TESOL Quarterly, 13, No. 2 (June 1979), pp. 219-227.

Bensoussan, Marsha. "A Comparison of Cloze and Multiple-Choice Reading Comprehension Tests," Report No. 57. Haifa: University Selection and Assessment Unit, June 1981.

..... "Manual for Teachers Preparing Examinations in English as a Foreign Language for Analysis by Computer," Report No. 44. Haifa: University Selection and Assessment Unit, Sept. 1979.

Cohen, Andrew. Testing Language Ability in the Classroom. Rowley, Mass.: Newbury House, 1980.

Cranney, A. Garr. "The Construction of Two Types of Cloze Reading Tests for College Students," Journal of Reading Behavior, 5, No. 1 (Winter 1972-73), pp. 60-64.

Greene, Frank P. "Modification of the Cloze Procedure and Changes in Reading Test Performances," Journal of Educational Measurement, 2, No. 2 (1965), pp. 213-217.

Guthrie, John T. "Reading Comprehension and Syntactic Responses in Good and Poor Readers," Journal of Educational Psychology, 65, No. 3 (1973), pp. 294-299.

Halliday, M.A.K. and Ruqaiya Hasan. Cohesion in English. Longman: London, 1976.

Jonz, Jon. "Improving on the Basic Egg: The M-C Cloze," Language Learning, 26, No. 2 (1976), pp. 255-265.

Klein-Braley, Christine. Empirical Investigations of Cloze Tests: An Examination of the Validity of Cloze Tests as Tests of General Language Proficiency in English for German University Students. Ph.D. dissertation, University of Duisburg, Germany: Liverpool, 1981.

Levenston, E.A., R. Nir, and S. Blum-Kulka. "Discourse Analysis and the Testing of Reading Comprehension by Cloze Techniques," paper presented at the International Symposium of Language for Special Purposes, Eindhoven, The Netherlands, August 2-4, 1982.

Nevo, B., E. Shor and R. Ramraz. "ITANA V, User's Guide: A Fortran Program for Multiple-Choice Tests and Item Analysis," Report No. 22. Haifa: University of Haifa Selection and Assessment Unit, Nov. 1975.

Oller, John Jr. "Scoring Methods and Difficulty Levels for Cloze Tests of Proficiency in English as a Second Language," Modern Language Journal, 56, No. 3 (March 1972), pp. 151-158.

..... "Cloze Tests of Second Language Proficiency and What They Measure," Language Learning, 23, No. 1 (1973), pp. 105-118.

..... and Christine A. Conrad, "The Cloze Technique and ESL Proficiency," Language Learning, 21, No. 2 (1971), pp. 183-195.

Ozete, Oscar. "The Cloze Procedure: A Modification," Foreign Language Annals, 10, No. 5 (Oct. 1977), pp. 565-568.

Pikulski, John J., and Edna C. Pikulski. "Cloze, maze, and teacher judgment," The Reading Teacher, 30, No. 7 (April 1977), pp. 766-770.

Porter, D. "Modified Cloze Procedure: A More Valid Reading Comprehension Test," English Language Teaching Journal, 30, No. 1 (Oct. 1975), pp. 151-155.

Ramraz, Rachel. "ITANA V: Computer Program for Item Analysis," Report No. 40. Haifa: University Section and Assessment Unit, June 1979.

Taylor, Wilson L. "Cloze Procedure: A New Tool for Measuring Readability," Journalism Quarterly, 30 (Fall 1953), pp. 415-433.

..... "Recent Developments in the Use of the Cloze Procedure," Journalism Quarterly, 33 (1956), pp. 42-48, 99.

Van Dijk, Teun A. Text and Context. Longman: London and N.Y., 1977.

..... Macrostructures. Lawrence Erlbaum: Hillsdale,N.Y., 1980.

APPENDIX A

Experiment Two

Fill-in-Test A

FILL-IN TEST A

POOR vs. RICH: A NEW GLOBAL CONFLICT

A conflict between two worlds - one rich, one poor is developing, and the battlefield is the globe itself. On one side are two dozen or so industrialized, non-Communist states _____(A)_____ 750 million citizens consume most of the world's resources, _____(B)_____ most of its manufactured goods, and enjoy history's _____(C)_____ standard of living. On the other side, demanding an ever larger share of that _____(D)_____, are about 100 underdeveloped poor _____(E)_____ with 2 billion people - millions of whom _____(F)_____ in the shadow of death by starvation or disease. _____(G)_____, the conflict has been limited to economic pressures and proposals, and _____(H)_____ in international forums. But the needs of the underprivileged nations are so pressing that some Western politicians describe _____(I)_____ as a "time bomb for the human race." The conflict _____(J)_____ the international economic system on which the _____(K)_____ of much of the world is based.

In the U.N. General Assembly, _____(L)_____ they are now a solid voting bloc, the developing states have approved resolutions that demand a "new international _____(M)_____ order." The meaning: massive and painful sacrifices by the rich _____(N)_____ the poor. So one-sided

APPENDIX A (continued)
Experiment Two
Fill-in Test A

FILL-IN TEST A

have the Assembly's actions become that the U.S. has _____(O)_____ them as "a tyranny of the majority."

_____(P)_____, the U.S. along with other First World nations, admits _____(Q)_____ there is a real grievance behind the angry rhetoric. Although the Third World population is literally exploding - there are 200,000 new mouths _____(R)_____ feed every day - the land available for growing food is _____(S)_____. As life in the countryside worsens, millions of peasants abandon their _____(T)_____ and head for the slums of the developing world's cities, vainly _____(U)_____ jobs that do not exist. Widespread poverty is a problem that afflicts all underdeveloped countries.

QUESTIONS ON FILL-IN TEXT A

	1	2	3	4
A.	what	whose	whom	which
B.	produce	but	for	occupation
C.	low	lowest	high	highest
D.	poverty	poor	wealth	rich
E.	people	states	living	industry
F.	to	great	exist	worried
G.	Finally,	However,	After which,	So far,
H.	speeches	industry	states	producing
I.	him	them	it	those
J.	must destroy	destroyed	could destroy	to destroy
K.	stabilized	stabilizer	stable	stability
L.	and	where	that	however
M.	economic	cultural	produce	politics
M.	in addition to	as a result of	on behalf of	in exchange for
O.	said	denounced	praised	told
P.	Furthermore,	Nonetheless,	Therefore,	Since,
Q.	in	by	to	that
R.	in	for	to	and
S.	diminishing	increasing	more fertile	less dry
T.	money	wives	towns	farms
U.	new	tried	seeking	finding

APPENDIX A

Experiment Two

Multiple-Choice Test A

MULTIPLE-CHOICE

TEXT A

A conflict between the worlds - one rich, one poor is developing, and the battlefield is the globe itself. On one side are two dozen or so industrialized, non-Communist states whose 750 million citizens consume most of the world's resources, produce most of its manufactured goods, and enjoy history's highest standard of living. On the other side, demanding an ever larger share of that wealth, are about 100 underdeveloped poor states with 2 billion people - millions of whom exist in the shadow of death by starvation or disease. So far, the conflict has been limited to economic pressures and proposals, and speeches in international forums. But the needs of the underprivileged nations are so pressing that some Western politicians describe them as a "time bomb for the human race." The conflict could destroy the international economic system on which the stability of much of the world is based.

In the U.N. General Assembly, where they now constitute a solid voting bloc, the developing states have approved resolutions that demand a "new international economic order." The meaning: massive and painful sacrifices by the rich on behalf of the poor.

So one-sided have the Assembly's actions become that the U.S. has denounced them as "a tyranny of the majority."

Nonetheless, the U.S., along with other First World nations, admits that there is a real grievance behind the angry rhetoric. Although the Third World population is literally exploding - there are 200,000 new mouths to feed every day - the land available for growing food is diminishing. As life in the countryside worsens, millions of peasants abandon their farms and head for the slums of the developing world's cities, vainly seeking jobs that do not exist. Widespread poverty is a problem that afflicts all underdeveloped countries.

QUESTIONS ON MULTIPLE-CHOICE TEXT A:

A. The aim of this passage is to
   1. explain a theory
   2. criticize a philosophy
   3. argue a cause
   4. present a problem

B. The industrialized, non-Communist sates (line 3) are the world's
   1. richest
   2. poorest
   3. more populated
   4. least developed

QUESTIONS ON MULTIPLE-CHOICE TEST A (continued)

C. or so (line 3). These words mean

1. either
2. about
3. in that way
4. therefore

D. its (Line 5). This word refers to

1. citizens
2. manufactured goods
3. the world
4. resources

E. According to lines 9-11, so far the underprivileged nations

1. have done very little to help themselves
2. have been in constant armed conflict
3. have no demands at all on the richer nations
4. have already put some pressure on the richer nations

F. Some Western politicians (line 12) think that the under-privileged nations

1. will have to wait until the richer nations can help them
2. will continue making speeches and political agreements in order to fulfill their needs
3. will use violence unless their needs are fulfilled
4. will have to realize that they will always remain poor

QUESTIONS ON MULTIPLE-CHOICE TEST A (continued)

G. According to lines 10-15, the conflict is dangerous because

1. it could upset the economic stability in the world
2. the poor nations might begin a nuclear war
3. the rich nations might begin a nuclear war
4. all the nations would become poor in the end

H. In the U.N. General Assembly, new resolutions demand that

1. the rich nations give much more to the poor nations
2. the rich nations approve of more poor nations
3. the poor nations sacrifice more for the rich nations
4. the poor nations make more massive efforts for the rich nations

I. According to lines 20-21, the U.S. thinks that the Assembly's resolutions

1. are fair to all nations
2. are in the interest only of the rich nations
3. are in the interest of only the poor nations
4. are insufficient in the fact of mass starvation

J. The word Nonetheless (line 22) is followed by a sentence which

1. agrees with the idea found in lines 20-21
2. contrasts with the idea found in lines 20-21
3. gives an example of the idea found in lines 20-21
4. explains the idea found in lines 20-21

QUESTIONS ON MULTIPLE-CHOICE

TEXT A (continued)

K. According to the author, the U.S. attitude towards the poor nations is that

1. they have a right to ask for help, although their language may be too violent
2. they have a right to ask for help, and their language is not violent enough
3. they have no right to ask for help, since they are only exaggerating the real situation
4. they have no right to ask for more help, since they are already getting the maximum that the First World is able to give

L. vainly (line 28). This word means

1. uselessly
2. proudly
3. conceitedly
4. irreverently

M. According to lines 22-30, peasants of the Third World Population who leave their farms

1. find a higher standard of living in the cities
2. develop the world's cities
3. diminish the land available for growing food
4. find no better life in the cities

QUESTIONS ON MULTIPLE-CHOICE

TEXT A (continued)

N. The author of the passage implies that

1. industrialization in a country probably leads to corruption
2. a decrease in population growth would help solve the problem of poverty
3. country life is much healthier than city life
4. the rich nations don't really want to help the poor nations

O. The main idea of this passage is that

1. the rich and the poor are in constant conflict
2. the industrialized non-Communist states are rich and powerful
3. poverty is a problem that must be solved by rich and poor nations alike
4. peasants are migrating from country to city in search of a better life

APPENDIX A

Table A: AN ANALYSIS OF THE TEST ITEMS IN MULTIPLE-CHOICE AND FILL-IN TESTS A

MULTIPLE-CHOICE TEST A

Number of Paragraph	Question	Function
1	A	Purpose of passage
	B	Paraphrase of sentence
	C	Phrase meaning
	D	Reference (within sentence)
	E	Paraphrase of sentence
	F	Paraphrase of phrase/sentence
	G	Paragraph meaning
2	H	Paragraph meaning
	I	Sentence meaning
3	J	Connecting word (meaning across sentence/paragraphs)
	K	Sentence meaning
	L	Word meaning
	M	Paragraph meaning
	N	Implication of passage
	O	Main idea of passage

Appendix A, Table A (continued)
(Fill-in Test A)

Fill-in Test A

Number of Paragraph	Question	Function
1	A	Relative pronoun, within sentence
	B	Verb, series, enumeration
	C	Comparative adjective, follow logical thought sequence
	D	Reference, across sentences, superordinate word
	E	Superordinate, ideas across sentences
	F	Verb, main idea of passage
	G	Connecting words, between sentences, follow line of argument
	H	Noun, enumeration within sentence
	I	Relative pronoun, reference within sentence
	J	Verb form, part of speech
	K	Noun form, part of speech
2	L	Relative pronoun, within sentence
	M	Adjective, main idea of sentence
	N	Prepositional phrase, main idea of paragraph/whole text
	O	Tone, choice of verbs/paragraph meaning

Appendix A, Table A (continued)
(Fill-in Test A)

Number of Paragraph	Question	Function
3	P	Connecting word, main idea, follow main argument in text
	Q	Collocation ("admit that there is ...")
	R	Collocation ("mouths to feed...") preposition
	S	Contrast -- fact, within sentence argument
	T	Contrast -- fact, within sentence argument
	U	Main idea of sentence

APPENDIX B: English Section of the Haifa University Entrance Examination*

1. SENTENCE COMPLETION --

DIRECTIONS: Select the item that best completes the sentence. There are four choices below each sentence. Write the number of the correct answer on the answer sheet.

EXAMPLE:

"What is that thing?" "That _____ a spider."

(1) to call (2) for calling (3) be called (4) is called

The correct English sentence is: "That is called a spider." (4)

Suggested time: 10 minutes.

1. "I'm not going to that movie." "_____ am I. I don't like musicals."

   (1) Too (2) Neither (3) Either (4) Also

2. "Who is the better musician?" "Betty is the _____ one."

   (1) more accomplished (2) accomplisher (3) more accomplish (4) accomplished more

3. "Why is Susan unhappy?" "Everyone went to the dance _____ she."

   (1) but (2) without (3) against (4) for

4. "That mother is not strict enough." "Yes, she lets her children _____ all over the magazine in the doctor's office."

   (1) writing (2) wrote (3) write (4) to write

* Sections I and II used with the kind permission of the English Language Institute, University of Michigan.

APPENDIX B (continued)

5. "Why are you worried?" "They _____got to hurry if they are going to catch their train."

   (1) have (2) are (3) must (4) will

6. John said he'll do it and he will because he's a _____boy.

   (1) confidential (2) reliable (3) trusting (4) permanent

7. "Is the treasurer in his office every day?" "Oh yes. _____ he must sign all checks, he must be here all the time."

   (1) Whether (2) However (3) Even (4) Since

8. "How do you know he was hungry?" "He stopped _____ to his girl friend in order to each lunch."

   (1) from writing (2) to writing (3) writing (4) to write

9. "Can't you make up your mind?" "No, they all look alike _____."

   (1) to me (2) from me (3) me (4) for me

10. "Do you watch television every night?" "No, but sometimes I wish I _____."

    (1) have time to (2) had time to (3) had time to do (4) have time to do

11. "What made you so angry with Bob's reply?" "_____ the stupid answers, Bob's took the prize."

    (1) All to (2) To all (3) All of (4) Of all

12. When no rain fell, the farmers feared a new _____.

    (1) tempest (2) invasion (3) plague (4) drought

APPENDIX B (continued)

II. VOCABULARY --

DIRECTIONS: Find the word that is closest in meaning to the underlined word in the sentence. Write the number of the correct answer on the sheet.

EXAMPLE:

It's too windy to go for a stroll.

(1) swim (2) sail (3) drive (4) walk

The word 'walk' means about the same thing as 'stroll' in this sentence. The sentence "It's too windy to go for a walk" (4), means the same thing as "It's too windy to go for a stroll." On the answer sheet under the sign *, you will find the number (4).

Suggested time: 10 minutes.

13. He purchased two cars.

(1) bought (2) owned (3) destroyed (4) sold

14. Why don't you tint it?

(1) iron (2) color (3) wave (4) smell

15. What ensued was interesting.

(1) remained (2) followed (3) increased (4) stopped

16. They roamed from town to town.

(1) walked (2) rode (3) drive (4) wandered

17. You shouldn't disclose that information.

(1) reveal (2) keep (3) hide (4) hold

18. The crest of the wave broke over the boat.

(1) force (2) bottom (3) top (4) weight

19. They hindered the sale of the newspaper.

(1) allowed (2) continued (3) helped (4) blocked

APPENDIX B (continued)

20. There was a weird light in the window.

    (1) bright (2) blue (3) strange (4) new

21. The magistrate forgot my name.

    (1) judge (2) minister (3) lawyer (4) chief

APPENDIX B

III M-C TEST --

DIRECTIONS: Below is a text with 5 choices for each question. Choose the items that best answers the question, and write its number on the answer sheet.

Ever since Hiroshima, it has been fashionable to say that another war would destroy civilization. Even Mr. Brezhnev fell into that phrasing once, though usually he makes it clear to his people that only capitalist civilization would be destroyed. But both the expressions of concern and the sometimes fantastic remedies that have been proposed to avert the danger have usually had a materialistic emphasis - as if civilization consisted of improved real estate, which would be flattened by hydrogen or atomic bombs. But civilization is not buildings, however beautiful or historic or whatever they contain. Civilization is something inside the people, or some of the people, who live and work in those buildings - the way they feel, the way they think, their capacity for thinking. Certainly it needs some economic foundation - more now than it used to, since now there must be some technological foundation too. But all that is only the background, not the thing itself.

-Elmer Davis, "Can Civilization Survive?"

APPENDIX B

22. The author of this passage thinks that
    1. civilization is created by the people
    2. civilization is mostly technical foundation
    3. another war will destroy civilization
    4. only capitalist civilization will be destroyed
    5. civilization consists of improved real estate

23. The author includes Mr. Brezhnev's statement to show that this attitude toward war is
    1. humourous
    2. a carelessly worded phrasing
    3. similar to a worldwide attitude
    4. unrealistic in view of the present worldwide situation
    5. a serious danger to civilization

24. The author objects to
    1. the buildings in civilization
    2. war and bombs
    3. Mr. Brezhnev's statement
    4. the materialistic definition of civilization
    5. the materialism of civilization

25. Civilization needs more economic foundation than before because
    1. it is wealthier
    2. it is technically more advanced
    3. it has more materialistic background
    4. it has more different kinds of money
    5. it contains buildings

FILL-IN

APPENDIX B

26. It has been fashionable to say (line 1) means

1. it is true
2. it has been proven true
3. it can be said
4. the fashion has been
5. people have said

27. though (line 3) means

1. even if
2. so
3. and
4. thus
5. through

28. concern (line 5) means

1. importance
2. worry
3. belonging
4. relation
5. interest

29. however (line 9) means

1. therefore
2. moreover
3. nevertheless
4. no matter how
5. but

APPENDIX B

30. since (line 14) means

    1. ago
    2. from the time that
    3. because
    4. even if
    5. subsequently

31. the thing itself (line 15) refers to

    1. the capacity for thinking
    2. civilization
    3. technical foundation
    4. economic foundation
    5. the background

IV. CLOZE PASSAGE --

DIRECTIONS: On the last page of the answer sheet you will find a text with missing items. For each blank space fill in one work, in English, to complete the thought of the sentence.

APPENDIX B

CLOZE PASSAGE

It is essential to give the child self-confidence in human nature through having faith in himself. There are a great many practices and _____ that operate against this seemingly simple goal. _____ is the traditional belief that human nature _____ basically evil, and that the primary function _____ education is to exercise the negative impulses _____ find ways to lead the child to perform functions in adult life. No matter where _____ stems from the concept that human nature _____ basically evil is likely to lead to practices that involve fear of pain and threats _____ punishment as techniques for bringing _____ human being. When this happens we are trying to create in a child a life-long _____ of guilt, a defeatist attitude about the world and an unhappy and unsatisfactory attitude about _____ as a person.

In place of this _____ we need to establish the point of _____ that each child is born with great _____ for positive living. Not only that he can inevitably become a positive, socially functioning organism, _____ that human nature has potentialities that can _____ shaped for good or evil according to _____ nature of experience and education.

APPENDIX B

CLOZE (2)

There is, _____course, the danger that the school will oppose society's demands, not taking the child's capacity into account. The school might also encourage high _____, which may cause chronic frustration in a _____who is unable to meet them. Education, _____, is not an automatic solution to the _____. Indeed, the school may be yet another _____ of difficulty for the child, creating obstacles rather than overcoming them.

Therefore, it is important _____the school to encourage the child's belief _____ the potentialities of human nature, and, by _____, his faith himself. If this goal is achieved, the child's school years and school experience will lead to a sense of accomplishment.

ראש היחידה: ר. רמרז
יועץ אקדמי: ד"ר ב. נבו
עוזר לראש היחידה: ק. פייזר
מזכירה: ס. עבאס, עוזרת למזכירה: ד. פולינס
מתכנתת: ח. לנדאו
פסיכומטריקאים: י. בורנשטיין, ג'. ספז, ש. לפיד
עוזרי מחקר: ע. יודפת, ח. אהרונסון, ש. ישראלי
כתובת: אוניברסיטת חיפה, הר הכרמל, חדר 784א', טלפון: 04-240336

דוחו"ת קודמים

45. השוואת מיועצים למועמדי שנה"ל תשל"ט.
46. תוקף הניבוי של מכשירי המיון בחוגי אוניברסיטת חיפה, בשנה"ל תשל"ח
47. נקודות חתך ומילוי "מיכסות חוגים" - תש"מ.
48. סולם מוקד השליטה של ROTTER (I-E Scale) ותיאור הסולם למידע פסיכומטרי על תרגומו העברי.
49. מועמדים לחוגים באוניברסיטת חיפה בשנים תשל"ד - תשל"ט שכיחות,מין,דת,גיל, ציוני בגרות,השגים במבחן הפסיכומטרי.
50. תלמידי שנה א' בחוגי אוניברסיטת חיפה בשנים תשל"ח - תש"מ שכיחות,מין,דת,גיל, ציוני בגרות, השגים במבחן הפסיכומטרי.
51. תהליך בניית מבחני הכניסה הפסיכומטריים לאוניברסיטת חיפה.
52. The Fill-in Test: combining multiple-choice and modified Cloze techniques to test English as a Foreign Language reading comprehension.
53. נקודות חתך חוגיות ומילוי מכסות חוגים תשמ"א.
54. לוח נורמות למבחן פיבודי.
55. תוקף ניבוי להצלחה בחוג לאנגלית.
56. Testing the test of advanced EFL reading comprehension: To what extent does the difficulty of a multiple-choice comprehension test reflect the difficulty of the text?
57. A Comparison of Cloze and Multiple-Choice Reading Comprehension Tests of English as a Foreign Language.
58. מועמדים לחוגים באוניברסיטת חיפה בשנים תשל"ה - תש"מ שכיחות,מין,דת,גיל, ציוני בגרות,השגים במבחני כושר שכלי.
59. תוקף הניבוי של מכשירי המיון בחוגי אוניברסיטת חיפה בשנת תשל"ט.
60. ניתוחי גורמים של סוללת משתני המיון תש"מ - תשמ"א.
61. Gender differences in cognitive functioning among university applicants in Israel: 1978-1980.
62. נקודות חתך חוגיות ומילוי "מכסות חוגים" תשמ"ב - המשך לדו"ח 53.
63. השפעת שונות ציונים תוך-נבדקית על תוקף הניבוי של סוללת מבחנים פסיכומטריים.